

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of:

Yasuhiro KAWAKITA, et al

Application. No.: *To Be Assigned*

Filed: *Concurrently*

FOR: ELECTRONIC DOCUMENT
INFORMATION EXPANSION APPARATUS,
ELECTRONIC DOCUMENT INFORMATION
EXPANSION METHOD, ELECTRONIC
DOCUMENT INFORMATION EXPANSION
PROGRAM, AND RECORDING MEDIUM
WHICH RECORDS ELECTRONIC
DOCUMENT INFORMATION EXPANSION
PROGRAM

Art Unit: *Unassigned*

Examiner: *Unassigned*

Atty. Docket No. 31759-190542

Customer No.



26694

PATENT TRADEMARK OFFICE

June 26, 2003

SUBMISSION OF PRIORITY DOCUMENTS

Commissioner for Patents
P.O. Box 1450
Alexandria, Virginia 22313-1450

Sir:

Submitted herewith are certified copies of Application No. JP2002-187695 filed on June 27, 2002 in Japan, and JP2003-002978 filed on January 9, 2003 in Japan, the priorities of which are claimed in the present application under the provisions of 35 U.S.C. 119. It is requested that the Examiner acknowledge receipt of the enclosed documents in the initial Office Action.

Respectfully submitted,


Norman N. Kunitz

Registration No. 20,586

VENABLE LLP

P.O. Box 34385

Washington, D.C. 20043-9998

Telephone: (202) 962-4800

Telefax: 202 (962-8300

NNK/elw
#466689

日 本 国 特 許 庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出 願 年 月 日

Date of Application:

2002年 6月27日

出 願 番 号

Application Number:

特願2002-187695

[ST.10/C]:

[J P 2 0 0 2 - 1 8 7 6 9 5]

出 願 人

Applicant(s):

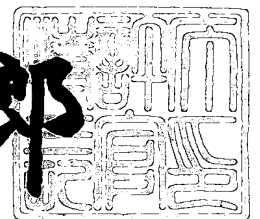
沖電気工業株式会社

Inventor: Yasuhiro KAWAKITA
Att'y Dkt: 31759-190542
Customer No: 26694

2003年 4月 1日

特 許 庁 長 官
Commissioner,
Japan Patent Office

太田 信一郎



出証番号 出証特2003-3022442

【書類名】 特許願

【整理番号】 KN002525

【提出日】 平成14年 6月27日

【あて先】 特許庁長官 及川 耕造 殿

【国際特許分類】 G06F 15/00

【発明者】

 【住所又は居所】 東京都港区虎ノ門1丁目7番12号 沖電気工業株式会社
社内

 【氏名】 川北 泰広

【発明者】

 【住所又は居所】 東京都港区虎ノ門1丁目7番12号 沖電気工業株式会社
社内

 【氏名】 池野 篤司

【特許出願人】

 【識別番号】 000000295

 【氏名又は名称】 沖電気工業株式会社

 【代表者】 篠塚 勝正

【代理人】

 【識別番号】 100090620

 【弁理士】

 【氏名又は名称】 工藤 宜幸

【手数料の表示】

 【予納台帳番号】 013664

 【納付金額】 21,000円

【提出物件の目録】

 【物件名】 明細書 1

 【物件名】 図面 1

 【物件名】 要約書 1

 【包括委任状番号】 9006358

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 電子メール文書情報拡充装置

【特許請求の範囲】

【請求項 1】 電子メール文書から、必要な要素で満たされた構造化データを生成する電子メール文書情報拡充装置であって、

電子メール文書を入力する入力手段と、

入力した電子メール文書から、所在情報を抽出する所在情報抽出手段と、

抽出された所在情報に基づき、必要なデータを取得する必要データ取得手段と

取得したデータを用いて、必要な要素で満たされた構造化データを生成する構造化データ生成手段と

を備えることを特徴とする電子メール文書情報拡充装置。

【請求項 2】 電子メール文書から、必要な要素で満たされた構造化データを生成する電子メール文書情報拡充装置であって、

電子メール文書を入力する入力手段と、

電子メール文書の内容を解析して情報単位に分割する分割手段と、

分割された個々の情報単位から、所在情報を抽出する所在情報抽出手段と、

抽出された所在情報に基づき、必要なデータを取得する必要データ取得手段と

取得したデータを用いて、必要な要素で満たされた構造化データを生成する構造化データ生成手段と

を備えることを特徴とする電子メール文書情報拡充装置。

【請求項 3】 上記必要データ取得手段は、抽出された所在情報に基づき、必要なデータを取得する際、所在情報が直接示す場所のデータと、情報単位に分割された部分の内容とが類似していることを判定した上で取得することを特徴とする請求項 1 又は 2 に記載の電子メール文書情報拡充装置。

【請求項 4】 上記必要データ取得手段は、抽出された所在情報に基づき、必要なデータを取得する際、所在情報が直接示す場所のデータと、さらにその場

所からたどることができるリンク先の場所のデータと、情報単位に分割された部分の内容とが類似していることを判定した上で取得することを特徴とする請求項 1 又は 2 に記載の電子メール文書情報拡充装置。

【請求項 5】 上記必要データ取得手段は、いずれかの所在情報が示す場所のデータと、情報単位に分割された部分とが、類似していることを判定する際、類似度がある一定値を超えた場合に、対応する所在情報が示す場所のデータを必要なデータとすることを特徴とする請求項 3 又は 4 に記載の電子メール文書情報拡充装置。

【請求項 6】 上記必要データ取得手段は、情報単位内で示されている所在情報が示す場所から、あらかじめ定められた階層までリンク先をたどり、全ての所在情報が示す場所のデータと、情報単位に分割された部分との類似度を算出し、最も類似度の高い対応する所在情報が示す場所のデータを必要なデータとすることを特徴とする請求項 3 又は 4 に記載の電子メール文書情報拡充装置。

【請求項 7】 上記必要データ取得手段は、必要データの取得のために行う類似度の判定を行う前に、取得したデータに対して、ハイパーリンク以外の制御文字を取り除く第 1 の前処理を行うことを特徴とする請求項 1 ～ 6 のいずれかに記載の電子メール文書情報拡充装置。

【請求項 8】 上記必要データ取得手段は、類似度の判定を行う前に、取得したデータに対して、主要箇所を抽出する第 2 の前処理を行うことを特徴とする請求項 1 ～ 7 のいずれかに記載の電子メール文書情報拡充装置。

【請求項 9】 上記構造化データ生成手段は、構造化データを生成する際、取得したデータからキーワードを抽出して、情報単位の内容と共に、構造化データを生成することを特徴とする請求項 1 ～ 8 のいずれかに記載の電子メール文書情報拡充装置。

【請求項 10】 上記構造化データ生成手段は、情報単位の内容から抽出したキーワードと、取得したデータから抽出したキーワードとを区別して、構造化データを生成することを特徴とする請求項 1 ～ 9 のいずれかに記載の電子メール文書情報拡充装置。

【発明の詳細な説明】

【 0 0 0 1 】

【発明の属する技術分野】

本発明は、電子メール文書から、必要な要素で満たされた構造化データを生成する電子メール文書情報拡充装置に関し、例えば、電子メール文書を情報源として扱う情報マネジメントシステムに適用し得るものである。

【 0 0 0 2 】

【従来の技術】

近年、電子メールの文書中に、情報の所在（URL）を記述して送信することが日常的に行われるようになってきている。電子メール閲覧ソフトはそれに対応して、そのURLを選択するだけで、Webブラウザソフトが起動するように連携するなどさまざまな工夫がされてきている。ところが、電子メールを着信した時点では、まだURLで示される場所にある情報は取得しておらず、利用者が改めてその情報を取得する動作を行う必要がある。

【 0 0 0 3 】

この問題点に着目して、特開 2 0 0 1 - 1 8 4 2 7 7 号公報では、電子メールの中に参照すべき情報の所在がURLで示されていると、そのURLが示す場所にある情報（HTML文書など）を自動的に取得し、受信した電子メールと関連付けて保存しておくという方法が紹介されている。これにより、電子メールを受信したユーザは、ネットワークとの接続が切れた状態でも、電子メール文書中のURLを指定するだけで、既に取得済みのデータを専用の表示装置で閲覧することができる。

【 0 0 0 4 】

【発明が解決しようとする課題】

上記公報によると、電子メール文書内に含まれるURLのデータを、電子メールと関連付けて全て取得するため、実際には、電子メール文書の内容とは関係のない部分まで取得することになる。したがって、表示装置内に全てのURLのデータを蓄積するのは、ネットワークとの接続が切れた状態でも、URLのデータを閲覧できる利点がある反面、蓄積効率が悪くなるという課題がある。

【 0 0 0 5 】

さらに、企業のURLが示されている時には、企業のトップページがリンクされていることが多く、このトップページのデータが蓄積されていると、前記専用の表示装置を利用して、電子メール文書の内容に関連する情報を探さなければならない。上記公報によると、指定したURLのページを取得して蓄積しているため、ネットワークの接続が切れた状態では、さらなるリンク先を探すことができないという課題が残されている。

【0006】

また、1つの電子メール文書の量が少ない場合には、十分なキーワードとのマッチングができず、必要な電子メールを的確に取得できないという問題もある。

【0007】

そのため、電子メール文書から、必要な要素で満たされた構造化データを生成できるような装置が望まれている。

【0008】

【課題を解決するための手段】

第1の本発明は、電子メール文書から、必要な要素で満たされた構造化データを生成する電子メール文書情報拡充装置であって、電子メール文書を入力する入力手段と、入力した電子メール文書から、所在情報を抽出する所在情報抽出手段と、抽出された所在情報に基づき、必要なデータを取得する必要データ取得手段と、取得したデータを用いて、必要な要素で満たされた構造化データを生成する構造化データ生成手段とを備えることを特徴とする。

【0009】

第2の本発明は、電子メール文書から、必要な要素で満たされた構造化データを生成する電子メール文書情報拡充装置であって、電子メール文書を入力する入力手段と、電子メール文書の内容を解析して情報単位に分割する分割手段と、分割された個々の情報単位から、所在情報を抽出する所在情報抽出手段と、抽出された所在情報に基づき、必要なデータを取得する必要データ取得手段と、取得したデータを用いて、必要な要素で満たされた構造化データを生成する構造化データ生成手段とを備えることを特徴とする。

【0010】

【発明の実施の形態】

(A) 実施形態

以下、本発明による電子メール文書情報拡充装置の一実施形態を図面を参照しながら詳述する。

【0011】

この実施形態は、図3に示すような電子メール文書に対して、URLで示された情報源にアクセスし、個々の情報と関連する内容を情報源から取得した上でキーワード抽出を行い、その結果を含めて構造化データを生成するものである。

【0012】

(A-1) 実施形態の構成

図1は、この実施形態の電子メール文書情報拡充装置の機能的構成を示すブロック図である。例えば、実施形態の電子メール文書情報拡充装置は、通信機能を有するパソコン等の情報処理装置で実現されるが、機能的には、図1で表すことができる。

【0013】

この実施形態の電子メール文書情報拡充装置は、入力部100、情報解析部101、外部データ取得部102、情報補完部（情報補充部）103および構造化データ生成部104を有する。

【0014】

入力部100は、情報および、その情報の情報源を示したURLを含む電子メール文書の入力を行うものである。

【0015】

情報解析部101は、入力された電子メール文書を、個々の情報単位に分割し、それぞれの情報単位から、情報源を示すURLを抽出するものである。

【0016】

外部データ取得部102は、情報解析部101で分割された情報単位毎に、そこに含まれているURLに基づいて、情報単位内に記載されている内容と類似する詳細なデータを外部の情報源から取得するものである。

【0017】

情報補完部 1 0 3 は、外部データ取得部 1 0 2 で取得したデータに対して、キーワードの抽出と主要箇所の抽出を行い、補充データを作成するものである。

【 0 0 1 8 】

構造化データ生成部 1 0 4 は、補充すべきデータをまとめて、構造化データを生成するものである。

【 0 0 1 9 】

入力部 1 0 0 は、電子メール文書を入力する機能を備えている。

【 0 0 2 0 】

情報解析部 1 0 1 は、入力した電子メール文書の内容を解析し、その中に含まれている複数の情報を個々の情報の単位に分割する機能を備えている。例えば、電子メール文書がニュースのメールマガジンであれば、1つの記事を1単位とする情報に分割する。そして、それぞれの情報単位の中に含まれるURLを抽出する機能を備えている。

【 0 0 2 1 】

外部データ取得部 1 0 2 は、URLなどで示される情報源から必要なデータを取得する。必要なデータの判断は、情報単位内に記載されている元の文章と、URLなどで示される情報源から取得したデータとの類似度を見ることで行う。

【 0 0 2 2 】

情報補完部 1 0 3 は、外部データ取得部 1 0 2 により取得したデータに対して、キーワード抽出および主要部の抽出を行い、元の情報単位に追加するデータを準備する。

【 0 0 2 3 】

構造化データ生成部 1 0 4 は、情報補完部 1 0 3 がそろえたデータを元に、構造化データの作成を行う。

【 0 0 2 4 】

(A-2) 実施形態の動作

図 2 は、実施形態の電子メール文書情報拡充装置の全体動作を示すフローチャートである。

【 0 0 2 5 】

ここでは、情報単位の例として、タイトル<TITLE>、概要<BODY>、キーワード<KEYWORD>、情報源の場所<URL>を、情報単位を構成する必須の内容とし、これらを全て含む構造化データの生成について説明する。そして、キーワードは全ての場合に生成されることになるが、電子メール文書に対して分割処理を行った後、概要が欠けている場合について以降で詳細に説明する。

【 0 0 2 6 】

ステップ S 2 0 0 の入力処理では、入力部 1 0 0 が電子メール文書を入力する。

【 0 0 2 7 】

ステップ S 2 0 1 の情報単位抽出処理では、情報解析部 1 0 1 が、入力された電子メール文書の中に含まれている情報を、それぞれ関連する内容毎に切り分ける。例えば、電子メール文書が、図 3 に示されているような場合には、図 4 に示すような情報単位に切り分けられる。このとき、切り分ける方法は、セパレータと呼ばれる特殊な記号の連続や、空白行などを手がかりに、これらで挟まれた部分を情報単位としても良い。また、段落や、表題記号などを手がかりにして、次の段落や表題記号が出現するまでを 1 つの情報単位としても良い。

【 0 0 2 8 】

そして、切り分けられた情報単位の中に、その情報の詳細情報の所在を示す URL が記載されていれば抽出する。

【 0 0 2 9 】

この実施形態では、抽出した結果をタグでマーキングするという形で表現している。例えば、図 4 に示す情報単位に対しては、図 5 に示すように抽出し、表現する。また、1 つの情報単位の中に、URL が複数個存在する場合にも、同様に抽出する。ただし、複数の出現については、URL の表現を区別するために、タグの中に i d という属性を付与し出力順に番号を付与している。URL の発見方法については、h t t p : / / で始まる文字列を検索するなど、一般的な方法を利用すればよい。また、URL 抽出後の表現方法も、複数の URL を識別できるのであれば、上記の方法に限るものではない。

【 0 0 3 0 】

ステップ S 2 0 2 のデータ取得処理（情報取得処理）では、外部データ取得部 1 0 2 が、ステップ S 2 0 1 で抽出した URL に基づいて、その URL が示す情報源からデータを取得する。これは一般に、ネットワークを介して URL が示すサーバにアクセスし、該当する HTML 文書を取得するという処理になる。

【 0 0 3 1 】

ステップ S 2 0 3 の判定処理では、データ取得処理 S 2 0 2 で取得した URL が示す場所のデータと、その URL が含まれていた情報単位の内容とが一致するかを判定する。判定の方法としては、例えば、取得したデータと、情報単位の内容のそれぞれからキーワードを抽出し、互いのキーワードの一致度を計算した結果、ある一定の値以上であれば一致と判定し、処理を S 2 0 5 に移す。そうでなければ不一致と判定し、処理を S 2 0 4 に移す。取得したデータに＜GET-DATA＞というタグをつけて表現した様子を図 6 に示す。

【 0 0 3 2 】

ここで、取得したデータは、一般的に HTML 文書と呼ばれる制御文字を含んだ文書である。そのため、あらかじめ取得したデータのハイパーリンク以外の制御文字を取り除く前処理を行った上で、判定処理を行っても良い。

【 0 0 3 3 】

また、取得したデータは、レイアウトによって記載内容を分類できるため、あらかじめ取得したデータの主要箇所を抽出する前処理を行った上で、抽出された主要箇所との判定処理を行っても良い。

【 0 0 3 4 】

ステップ S 2 0 3 の判定処理において、不一致と判定された場合の処理として、ステップ S 2 0 4 の URL 変更処理では、先に取得したデータの中に含まれるハイパーリンクを抽出して、第 1 階層の URL リストを作成した上で一時的に記憶しておき、各 URL に対してデータ取得処理 S 2 0 2 及び判定処理 S 2 0 3 を繰り返す。第 1 階層の URL リストに対して全て不一致となった場合は、一時的に記憶されている第 1 階層の URL リストから取得できるデータに対して、再びハイパーリンクを抽出して第 2 階層の URL リストを作成して一時的に記憶して

おき、各URLに対してデータ取得処理S 2 0 2及び判定処理S 2 0 3を繰り返す。

【 0 0 3 5 】

ここで、階層の深さは自由に設定でき、繰り返しの回数を制限できるものとする。そして、指定された階層に対して全ての取得データとの判定処理を行った結果、一致するものが無ければ、処理はS 2 0 5に移される。

【 0 0 3 6 】

ステップS 2 0 5のキーワード抽出処理は、情報補完部1 0 3の処理の一つで、情報単位の中に含まれる内容と、取得したデータに対して、キーワードとして扱う文字列の抽出を行う。抽出方法については、特に定めるものではなく従来の方法を用いてよい。ただし、情報単位の中に含まれるキーワードであるのか、取得したデータの中に含まれるキーワードであるのかは、情報単位に対する検索を行う際に、検索対象を選択可能にするため、区別して情報単位を管理する。

【 0 0 3 7 】

例えば、図7に示すように、情報単位の中と取得したデータからキーワードを抽出し、情報単位の中にタグをつけて表現しておく。このとき、キーワードタグの属性値を利用して、情報単位の中に含まれるものであれば、T（タイトル部分）、D（概要部分）、取得したデータからであれば、G、複数の部分に含まれているものであれば、その記号を列挙する。

【 0 0 3 8 】

ステップS 2 0 6の主要部抽出処理は、情報補完部1 0 3の処理の一つで、取得したデータの中から、主要部のみを抽出する。ここでの主要部とは、取得したデータの中から、情報単位が示す内容と類似する、または情報単位が示す内容の詳細に相当する部分をいう。主要部として抽出する文字数を制限しなければ、取得したデータを全て主要部として扱ってもよいが、具体的にはある文字数で制限し、その文字数に収まるように、取得したデータから主要部を抽出する。

【 0 0 3 9 】

例えば、図8に示すように、＜GET-DATA＞というタグで囲むことにより表現されている取得したデータから、主要部を抽出し、＜BODY＞というタ

グで囲んで、情報単位の中に表現する。このとき、この主要部が取得したデータから得られたものであることを示す情報として、属性値にGという記号をつけておく。情報単位内にもともと含まれていた場合には、属性値にOという記号をつけておく。

【0040】

ステップS207の構造化データ作成処理は、構造化データ生成部104の処理で、情報単位の内容と、キーワード抽出処理S204の結果と、主要部抽出処理S205の結果を組み合わせ、構造化データを生成する。例えば、図9に示すように、タグ付けした形で、構造化データを生成する。このとき、取得したデータについては、不要なものも含まれているため、主要部を抽出した後は削除することで、蓄積効率を上げる。

【0041】

ステップS208の判定処理は、情報単位抽出処理S201において、複数の情報単位が抽出された場合に、処理されていない情報単位がないか判定する。処理されていない情報単位があれば、ステップS202に処理を移す。全ての情報単位に対して処理されていれば、生成した全ての構造化データを出力する。

【0042】

(A-3) 実施形態の効果

上記実施形態に示した装置を、メールサーバ、もしくはメールクライアントの1機能として動作させることにより、電子メール文書の中に、URLで示されている部分があった場合、既にURLで示された部分から、電子メール文書の内容に該当するデータが読み込まれた状態で、電子メールを端末に受信できるため、ユーザは、改めてURLの指定や、URLの情報を取得する必要がなく、メールを受信した時点で必要な情報を得ることができる。

【0043】

また、取得したデータ全てを蓄積するのではなく、電子メール文書の内容に相当するデータの中から、さらに主要部のみを抽出して蓄積するため、蓄積効率が良い。

【0044】

また、電子メールの受信だけで、URLの情報も同時に取得できているので、電子メールの閲覧ソフトウェアだけで、必要なURL情報も見ることができる。

【0045】

また、タイトルとURLだけからなるような情報に対して、URLが示すサーバから取得したデータに対してキーワード抽出を行った上で、構造化データを作成するため、本構造化データをデータベース等に蓄積した上でキーワード検索をする際、検索精度が、タイトルだけに対して検索を行った場合に比べて、格段に向上する。

【0046】

(B) 他の実施形態

本装置の最終出力の形式は、必要に応じて電子メール文書の形態や、Webブラウザで閲覧可能な形態に変換してもよい。そして、これらを、電子メールとしてユーザに送信しても良い。

【0047】

また、情報単位の内容とURLが示すサーバから取得したデータとの類似度を判定する際、あらかじめ指定した階層までの全てのリンク先のデータを取得して類似度を計算した上で、最も類似度の高いデータを採用するようにしても良い。

【0048】

また、ステップS205のキーワード抽出処理とS206の主要部抽出処理の順序を入れ替えても良い。その場合は、主要部抽出処理の結果に対して、キーワード抽出処理を行うことになる。

【0049】

また、入力される電子メール文書が、複数の情報を含んでいない場合には、分割処理手段を備えなくてもよい。

【0050】

【発明の効果】

以上のように、本発明によれば、電子メール文書から、必要な要素で満たされた構造化データを生成できる電子メール文書情報拡充装置を提供することができる。

【図面の簡単な説明】

【図 1】

施形態の電子メール文書情報拡充装置の機能的構成を示すブロック図である。

【図 2】

実施形態の電子メール文書情報拡充装置の全体動作を示すフローチャートである。

【図 3】

電子メール文書の一例を示す説明図である。

【図 4】

実施形態の情報解析部による図 3 の文書に対する情報単位抽出処理の結果例を示す説明図である。

【図 5】

実施形態の情報解析部による抽出情報単位に対する URL の抽出結果例を示す説明図である。

【図 6】

実施形態の外部データ取得部の取得結果例を示す説明図である。

【図 7】

実施形態のキーワード抽出処理の処理結果例を示す説明図である。

【図 8】

実施形態の主要部抽出処理の処理結果例を示す説明図である。

【図 9】

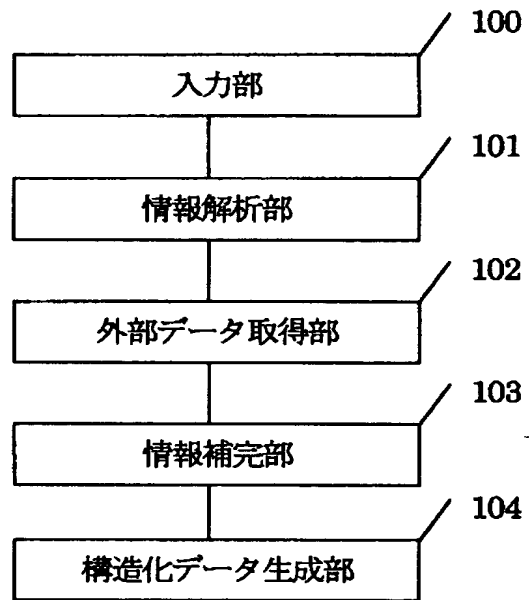
実施形態の構造化データ生成処理による構造化データ例を示す説明図である。

【符号の説明】

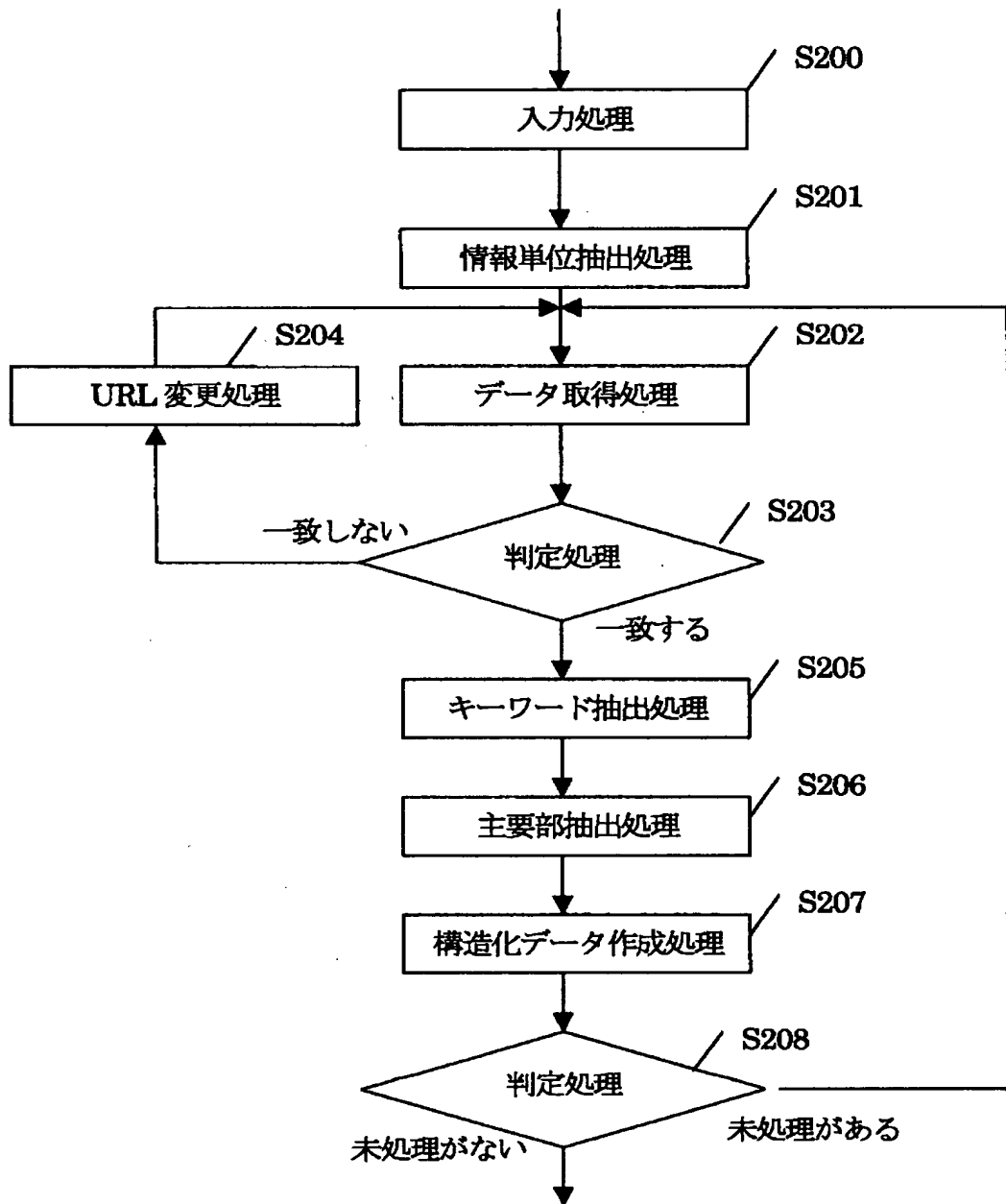
1 0 0 …入力部、1 0 1 …情報解析部、1 0 2 …外部データ取得部、1 0 3 …情報補完部（情報補充部）、1 0 4 …構造化データ生成部。

【書類名】 図面

【図 1】



【図 2】



【図 3】

<p>.....</p> <p>〇〇会社、△△を発表 〇月×日、〇〇会社は、新製品△△を発表した。価格は、2500 円～10000 円。全国の家電量販店等で〇月×日より発売を開始する。 http://www.xxx.zz/message1.htm</p> <p>.....</p> <p>××銀行と△△銀行が合併 http://www.yyy.zz/message2.htm</p> <p>.....</p> <p>情報系新会社△△研究所が誕生 http://www.zzz.zz/message3.htm http://www.aaa.zz/release1.htm</p> <p>.....</p> <p>⋮ ⋮ ⋮</p>
--

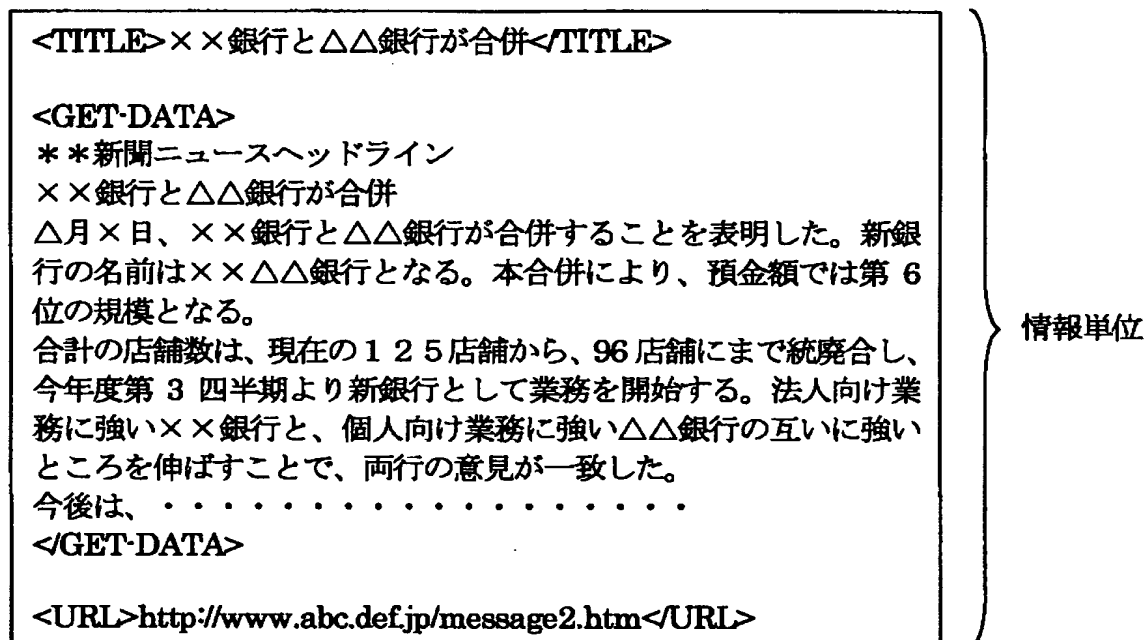
【図 4】

○○会社、△△を発表 ○月×日、○○会社は、新製品△△を発表した。価格は、2500 円 ～10000 円。全国の家電量販店等で○月×日より発売を開始する。 http://www.abc.def.jp/message1.htm	情報単位
××銀行と△△銀行が合併 http://www.abc.def.jp/message2.htm	情報単位
情報系新会社△△研究所が誕生 http://www.abc.def.jp/message3.html http://www.aaa.zz/release1.htm	情報単位

【図 5】

<TITLE>○○会社、△△を発表</TITLE> <BODY> ○月×日、○○会社は、新製品△△を発表した。価格は、2500 円 ～10000 円。全国の家電量販店等で○月×日より発売を開始する。 </BODY> <URL id=1> http://www.abc.def.jp/message1.htm </URL>
<TITLE>××銀行と△△銀行が合併</TITLE> <URL id=1> http://www.abc.def.jp/message2.htm </URL>
<TITLE>情報系新会社△△研究所が誕生</TITLE> <URL id=1> http://www.abc.def.jp/message3.html </URL> <URL id=2> http://www.aaa.zz/release1.htm </URL>

【図 6】



【図 7】

<TITLE>××銀行と△△銀行が合併</TITLE>

<GET-DATA>

**新聞ニュースヘッドライン

××銀行と△△銀行が合併

△月×日、××銀行と△△銀行が合併することを表明した。新銀行の名前は××△△銀行となる。本合併により、預金額では第 6 位の規模となる。

合計の店舗数は、現在の 125 店舗から、96 店舗にまで統廃合し、今年度第 3 四半期より新銀行として業務を開始する。法人向け業務に強い××銀行と、個人向け業務に強い△△銀行の互いに強いところを伸ばすことで、両行の意見が一致した。

今後は、・・・・・・・・・・・・・・・・

</GET-DATA>

<KEYWORDS>

<KEYWORD att=TG>××銀行</KEYWORD>

<KEYWORD att=TG>△△銀行</KEYWORD>

<KEYWORD att=G>××△△銀行</KEYWORD>

<KEYWORD att=G>第 6 位</KEYWORD>

<KEYWORD att=G>125 店舗 </KEYWORD>

<KEYWORD att=G>96 店舗 </KEYWORD>

⋮

⋮

<KEYWORDS>

<URL>http://www.abc.def.jp/message2.htm</URL>

情報単位

【図 8】

```

<TITLE>××銀行と△△銀行が合併</TITLE>
<BODY att=G>
△月×日、××銀行と△△銀行が合併することを表明した。新銀行の名前は××△△銀行となる。本合併により、預金額では第 6 位の規模となる。
</BODY>
<GET-DATA>
**新聞ニュースヘッドライン
××銀行と△△銀行が合併
△月×日、××銀行と△△銀行が合併することを表明した。新銀行の名前は××△△銀行となる。本合併により、預金額では第 6 位の規模となる。
合計の店舗数は、現在の 1 2 5 店舗から、96 店舗にまで統廃合し、今年度第 3 四半期より新銀行として業務を開始する。法人向け業務に強い××銀行と、個人向け業務に強い△△銀行の互いに強いところを伸ばすことで、両行の意見が一致した。
今後は、・・・・・・・・・・・・・・・・
</GET-DATA>
<KEYWORDS>
  <KEYWORD att=TG>××銀行</KEYWORD>
  <KEYWORD att=TG>△△銀行</KEYWORD>
  <KEYWORD att=G>××△△銀行</KEYWORD>
  <KEYWORD att=G>第 6 位</KEYWORD>
  <KEYWORD att=G>125 店舗 </KEYWORD>
  <KEYWORD att=G>96 店舗 </KEYWORD>
  :
  :
</KEYWORDS>
<URL>http://www.abc.def.jp/message2.htm</URL>

```

情報単位

【図 9】

```
<INFO>
<TITLE>××銀行と△△銀行が合併</TITLE>
<BODY att=G>
△月×日、××銀行と△△銀行が合併することを表明した。新銀行の
名前は××△△銀行となる。本合併により、預金額では第 6 位の規模となる。
</BODY>
<KEYWORDS>
  <KEYWORD att=TG>××銀行</KEYWORD>
  <KEYWORD att=TG>△△銀行</KEYWORD>
  <KEYWORD att=G>××△△銀行</KEYWORD>
  <KEYWORD att=G>第 6 位</KEYWORD>
  <KEYWORD att=G>125 店舗 </KEYWORD>
  <KEYWORD att=G>96 店舗 </KEYWORD>
  :
  :
</KEYWORDS>
<URL>http://www.abc.def.jp/message2.htm</URL>
</INFO>
```

情報単位

【書類名】 要約書

【要約】

【課題】 電子メール文書から、必要な要素で満たされた構造化データを生成できる電子メール文書情報拡充装置を提供する。

【解決手段】 本発明の電子メール文書情報拡充装置であって、電子メール文書を入力する入力手段と、入力した電子メール文書から、所在情報を抽出する所在情報抽出手段と、抽出された所在情報に基づき、必要なデータを取得する必要データ取得手段と、取得したデータを用いて、必要な要素で満たされた構造化データを生成する構造化データ生成手段とを備えることを特徴とする。ここで、電子メール文書の内容を解析して情報単位に分割する分割手段を設け、所在情報抽出手段が、分割された個々の情報単位から、所在情報を抽出するようにしても良い。

【選択図】 図 1

出 願 人 履 歴 情 報

識別番号 [000000295]

1. 変更年月日	1990年 8月22日
[変更理由]	新規登録
住 所	東京都港区虎ノ門1丁目7番12号
氏 名	沖電気工業株式会社